云 南 植 物 研 究 2006, **28** (2): 127~132
Acta Botanica Yunnanica

ITS 假基因对栎属系统学研究的影响及其对分子系统学研究的启示*

马长乐[1,2], 周浙昆[1**]
(1 中国科学院昆明植物研究所生物多样性与生物地理学重点实验室, 云南 昆明 650204;
2 中国科学院研究生院, 北京 100039)

摘要: 核糖体 nrDNA ITS 是被子植物系统发育研究中应用最广泛的分子标记之一。以前人们认为同一物种中的 ITS 序列因致同进化而使不同拷贝高度一致, 在分子系统学研究中常以 ITS1-5.8S-ITS2 序列作为构建系统进化树的基础。近年来, 在对一些被子植物的研究中发现这段序列在同一物种中具有多态性, 有些拷贝中的 5.8S 区不具编码功能, 人们把含有不具编码功能 5.8S 区的 ITS1-5.8S-ITS2 序列定义为 ITS 假基因序列, 它对同源基因致同进化的假设形成了新的挑战。在诸多应用 ITS 序列重建系统进化关系的研究中, 栎属系统学研究因 ITS 假基因的发现而倍受关注。本文以栎属为例回顾了 ITS 假基因的发现过程, 分析了其对该属系统学研究的影响, 为分子生物学在植物系统进化研究中的应用提供一些新的参考。

关键词: ITS; 假基因; 栎属; 分子系统学
中图分类号: Q 75, Q 949　　文献标识码: A　　文章编号: 0253-2700(2006)02-127-06

Effect of ITS Pseudogene on the Phylogenetic Study of *Quercus* (Fagaceae) and Its Revelation on the Plant Molecular Phylogenetics

MA Chang-Le[1,2], ZHOU Zhe-Kun[1**]
(1 *Laboratory of Biodiversity and Biogeography*, *Kunming Institute of Botany*, *Chinese Academy of Sciences*, Kunming 650204, China;
2 *Graduate School*, *Chinese Academy of Sciences*, Beijing 100039, China)

Abstract: Sequences of ITS in nrDNA is one of the widely used markers in angiosperm systematics. Many early reports of ITS region (ITS 1, 5.8S and ITS 2) variation in flowering plants indicated that nrDNA arrays within individuals are homogeneous. However, recent studies have discovered that individuals often contain potentially non-functional nrDNA copies (pseudogenes). These findings suggest that complete concerted evolution should not be assumed when embarking on phylogenetic studies using nrDNA sequences. ITS pseudogene has been detected in the genus of *Quercus* L. and misled the phylogenetic reconstruction, which made this genus frequently focused in the related studies. The authors take this genus as an example to discuss the influence of the pseudogene on the phylogenetic study, conclude that combining application of molecular evolution and classic taxonomy knowledge in plant systematics study may lead to more convincible results.

Key words: ITS; Pseudogene; *Quercus*; Molecular phylogenetics

植物系统学以植物各类群之间的演化关系为研究对象, 分子生物学与植物系统学相结合为植物系统学研究带来了一次巨大的变革 (Savolainen and Chase, 2003)。大量的分子标记成为了研究

---

* 基金项目: 国家自然科学基金 (30540077) 和吴征镒先生 2001 年云南省科学技术突出贡献奖资助自选项目 (KIB-WU-2001-01)

** 通讯作者: Author for correspondence. E-mail: zhouzk@mail.kib.ac.cn

收稿日期: 2005-01-17, 2005-09-30 接受发表

作者简介: 马长乐 (1976-) 男, 博士研究生, 主要从事植物地理学与生物多样性方面的研究。

生物系统演化关系的重要依据（邹喻苹等，2001），分子技术和一些新的分析方法的引入也为深入探索植物系统演化关系提供了可靠的基础（Soltis 等，2000）。在众多的分子片段中核糖体 rDNA 的转录间隔区（ITS）序列已经成为了当前被子植物系统学研究中最常用到的分子标记之一（Alvarez and Wendl，2003）。核糖体 rDNA 中的 18S、5.8S、28S 基因序列是中度重复序列，进化缓慢而且相对保守，但位于这 3 个基因序列之间的转录间隔区（ITS）序列却进化的相当迅速。以前，一般认为因致同进化同一物种 ITS 序列的不同拷贝高度一致（Bailey 等，2003），在分子生物学实验中又可以使用通用引物扩增得到 ITS1-5.8S-ITS2 的序列，所以研究人员常以这段序列作为构建进化树的基础。现在，ITS 序列已经成为了研究较低分类群（属间或种间）系统关系的重要标记（田欣和李德铢，2002）。

栎属（*Quercus* L.）是壳斗科中的一个大属，在北半球有着广泛的自然分布，具有重要的经济价值（Nixon，1993），依据不同的分类系统该属在全世界共有 300 种（陈焕镛和黄成就，1998；黄成就和张永田，1999）到 531 种（Govaerts and Frodin，1998），此外对栎属所包括的种数还有一些其它的观点（Lawrence，1951；Elias，1971）。栎属有悠久的化石历史，从第三纪起就是北半球森林生态系统中的优势种和建群种，已经成为了植物系统演化、地史和全球气候变迁中的一个重要研究对象（周浙昆，1992 a, b，1993，1999；Manos 等，1999，2001）。近年来，曾有人基于栎属植物叶器官形态结构特征进行分支分析（普春霞等，2002），但分子系统学方法在栎属系统演化研究中却发挥了重要的作用，解决了许多传统方法不易得到解答的问题。然而，在应用 ITS 序列作为分子标记对栎属的系统学研究中发现该属植物同一个体中的多个 ITS 序列存在差异，一部分是假基因序列。这一发现成为了应用 ITS 序列做进化分析的一个新的挑战，因为同源基因致同进化的假设就不再成立，相应的系统进化分析就会因同一物种中多样化的序列类型而无法构建出可靠的进化树（Bailey 等，2003）。因栎属系统学研究在 ITS 假基因发现过程中具有重要作用（Alvarez and Wendel，2003；Bailey 等，2003），本文以栎属为例回顾了 ITS 假基因的发现过程，为分子生物学在植物系统进化研究中的应用提供一些新的参考。

1 栎属中 ITS 假基因的发现

Samuel 等（1998）选择了 10 种栎属植物，以核糖体 rDNA 的 ITS 序列为标记对该属的系统进化关系进行了研究：根据已知的 18S/26S 序列设计引物（Q1/Q2）进行 PCR 扩增，PCR 产物胶回收纯化后又对 *Quercus ilex* 和 *Fagus sylvatica* 克隆测序，其它样品用 PCR 产物直接测序，测序获得的序列与已发表的 ITS 序列相比对以确定间隔区（ITS1/ITS2）与编码区（5.8S）。Manos 等（1999）应用 ITS 作为分子标记对 25 个种，包括 Samuel 等（1998）所用的 8 个种（表 1），对壳斗科的分子系统学进行了研究：用 ITS4/ITS5 作为引物，PCR 产物直接用TA 克隆试剂盒克隆，挑克隆并用引物 ITS4/ITS5 进行 PCR 扩增检验，在 Biosystems 373 测序仪上测序，测得的序列与已发表的被子植物 ITS 序列相比对，确定间隔区（ITS1/ITS2）与编码区（5.8S），还用 G/C 含量、序列长度、序列的变异度以及信息位点的数量作为依据确定所获序列为目的序列。两个研究组都获得了包括 ITS1、5.8S、ITS2 的序列，并用相同的方法构建系统进化树，分析得到的系统关系却有很大的不同。Manos 等（1999）认为造成这一结果的原因是他们所用的技术方法不同。

表 **1** **Samuel（1998）**取样列表

Table 1 Sampled species in the study of Samuel (1998)

Outgroup	*Fagus sylvaica* L.*	*Castanea sativa* Miller*
Ingroup	*Q. suber* L. *Q. acutissima* Carr. *Q. virginiana* Miller *Q. coccifera* L. *Q. robur* L.	*Q. macrolepis* Kotschy* *Q. rubra* L. *Q. ilex* L. *Q. cerris* L. *Q. petraea* (Matt). Liiebl.*

标有* 的种类为仅 Samuel (1998) 中有，而在 Manos (1999) 没有。

究竟是什么原因造成这样的不同呢？Mayol and Rossello（2001）对这两个研究组所得的序列的长度、G+ C 含量、应用 MFOLD（versions 2.3 和 3.0）软件构建的RNA 的二级结构的热动力学稳定性进行了分析，发现 Samuel 等（1998）测得的序列GC 含量变异大，在编码区（5.8S）长度有变化，碱基替代率高，既使在高保守区情况也

是如此，RNA 二级结构稳定性较低，表现出种间差异大的特点，认为他得到的是假基因序列。而Manos 等（1999）获得的序列经分析认为是功能基因序列。Mayol and Rossello（2001）认为造成两者获得的序列不同的原因并不是因为不同的实验方法造成的，而是Samuel 等（1998）试验得到的是一个没有功能的平行进化的同源基因片段，即含有假基因的序列，应用其得到的系统进化分析结果不可靠。

Muir and Sch┣tterer（1999）对一份 *Q. petraea*

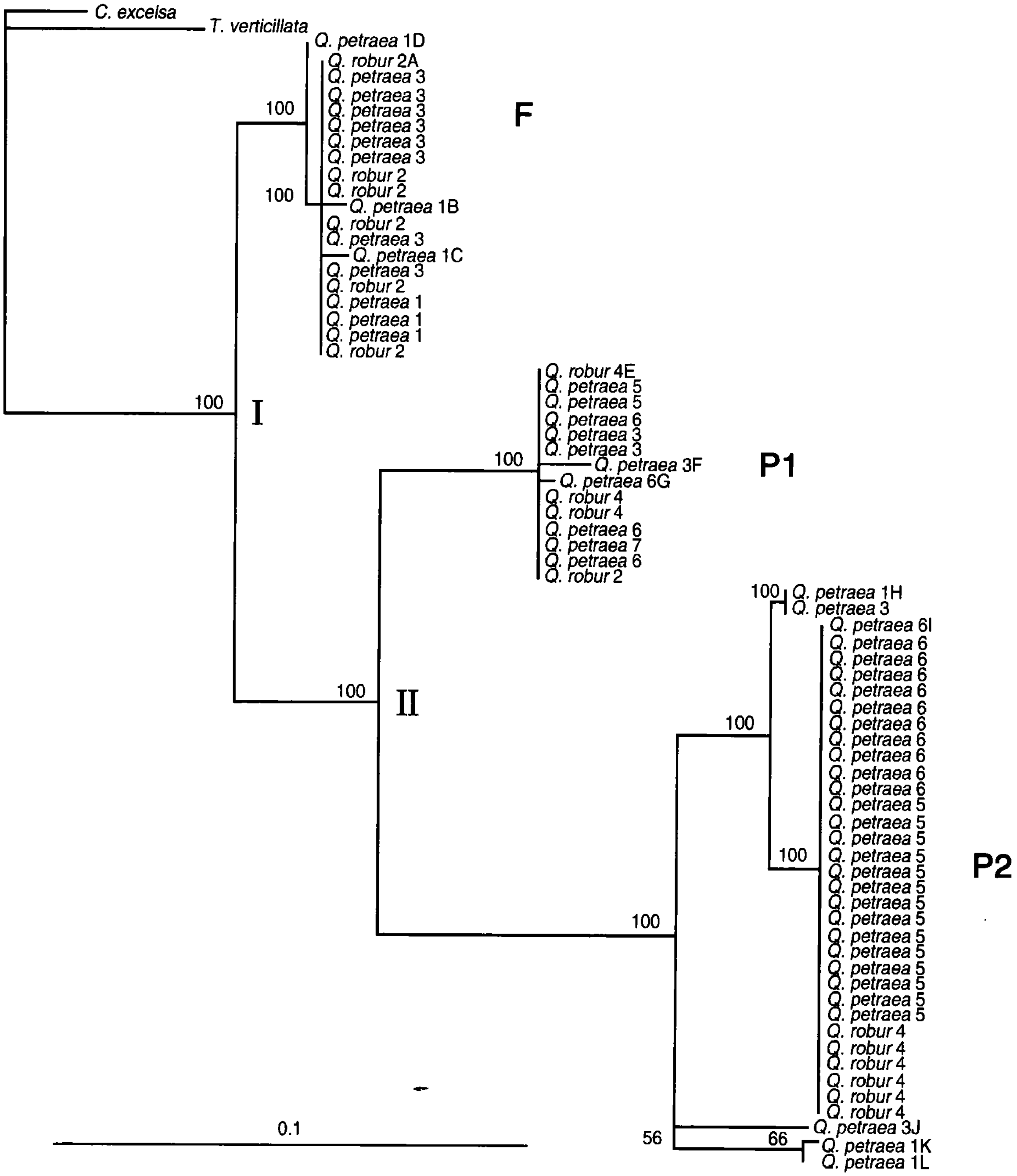

图1　*Quercus petraea* 和 *Quercus robur* 的 70个不同的克隆的 ITS2 和 5.8S 序列构建的最大似然树（Maximum-likelihood tree）；以 *Colombobalanus excelsa* 和 *Trigonobalanus verticillata* 作为外类群（引自 Muir 等，2001）

Fig. 1　Maximum-likelihood tree of 70 combined 5.8S and ITS2 PCR clones from *Quercus petraea* and *Quercus robur*, *Colombobalanus excelsa* and *Trigonobalanus verticillata* as the outgroup

材料用PCR-clone的方法实验得到了5.8S（161 bp）和ITS2（228 bp）序列，经过序列分析，从同一PCR产物的13个克隆中得到的序列可以分为3种类型，这就意味着在核仁组织区（Nucleolus Organizer Regions，NORs）有3种类型的ITS序列存在。在此之前Zoldos等（1999）用荧光原位杂交的方法确定了*Q.petraea*的NORs分布于不同的染色体上。Muir等（2001）用在自然界中常出现杂交的*Q.petraea*和*Q.robur*来研究nrDNA的演化，发现这2个种的ITS2+5.8S序列都可以分为3种类型（图1），也就是一个具有3种基因类型的nrDNA家族，其中只有1种序列类型的5.8S区适用于系统进化分析。

由此可见，无论是克隆测序或PCR产物测序，不同的研究者实验所得的相同物种的基因序列因为有假基因的存在才出现差异，因为假基因与功能基因的进化速率不同，不能反映真实的进化速率（至少是现在分子生物学的研究观点），所以用假基因或包含有假基因的序列为依据构建的系统进化树并不可靠，这也正是Samuel等（1998）与Manos等（1999）分析结果不同的原因。

2 ITS假基因的特性

在生命体漫长的演化进程中，假基因不断的产生、逐渐积累，大量的假基因存在于动物、植物、微生物中，一般认为它们在进化上与正常的功能基因相互关联。在基因的复制过程中，基因序列发生着不同形式的变化，包括突变、片段的漂移、插入、缺失等，如果基因序列发生了剧烈的变化而失去编码功能，就成为假基因。大约80%的假基因源自于逆转录，另有20%源自于基因的复制（ISCID，2004）。

因为ITS1-5.8S-ITS2在一个或多个染色体位点上的多个重复，rDNA上基因位点在进化上易发生变化，造成了整个进化过程中并不是所有的重复片段都具有功能，其中的一部分退化成为假基因（Makalowski，2000）。这些假基因并不会在短时间内从基因组中消失，虽然它们失去了转录的功能，却在多位点的致同进化中发生着变化。这些假基因会存在于基因组的某个片段中，或许它们也以不同于功能基因的进化速率发生着独立的变化（Bailey等，2003）。nrDNA中ITS假基因已在被子植物的不同类群都有发现（Buckler等，1996；Yang等，1999；Kita and Ito，2000；Mayol and Rossello，2001；Hartmannt等；Muir等，2001）。在对不同植物类群的ITS假基因与功能基因对比后发现它们的二级结构稳定性较低，因去氨基过程导致AT含量增高，在保守区内的碱基替代率相对较高（Buckler等，1996；Alvarez and Wendel，2003），Bailey等（2003）已经对假基因的鉴别依据与方法作了详细介绍，本文就不再赘述了。

3 ITS序列在栎属系统学研究中的应用

尽管栎属ITS序列存在假基因现象，我们仍然可以依据假基因的特性，通过一定的实验设计得到含有具有编码功能的5.8S区的ITS1-5.8S-ITS2序列。Manos等（1999）综合了叶绿体DNA和ITS序列的数据，所得出的分支树清晰地分出了4个单系类群：（*Cerris*-（*Lobatae*-（*Protobalanus* + *Quercus sensu stricto*））），这一结果与Nixon（1993）依据形态数据的对栎属内的分类观点比较，其中的3个组的聚类是很吻合的，只有*Cerris* group”和“*Ilex* group”处于同一分支上而被合并为Sect. *Cerris*。Manos and Standord（2001）在对壳斗科做系统进化研究时，为避免假基因，在选择ITS序列时遵从了以下的3个原则：（1）间隔区（ITS1/ITS2）的差异最小和5.8S区的最大的保守性；（2）分支内部分异平均水平以及所有样本的平均分异水平，即每个分支内各序列的分异以及所有样本的分异水平在一个适度的范围内；（3）经典的系统分类学知识，同时还应用BLAST在Genebank中与其它的ITS进行了比对，将栎属ITS序列提供的系统学信息与化石资料相结合，应用DIVA（dispersal-vicariance analysis）分析对栎属的分布模式研究表明：栎属明显地分化成为两支，并与它们的地理分布相对应，即新世界分布和旧世界分布，并且这两支在很早以前就发生了间断，随后分化成不同的种类，其分布局限于中纬度地区。他的研究结果与早先Axelrod（1983）认为栎属中的主要类群是原衍生地独立演化的观点相一致。

4 ITS假基因对分子系统学研究的启示

假基因在植物和动物分子系统学研究领域常

用的 nrDNA (ITS) 和 mtDNA (Zhang and Hewitt, 2003) 中的存在对分子进化研究产生了很大的影响，同时也给分子系统学研究带来了新的启示。一方面，随着分子生物学的深入发展，人们通过测序得到越来越多的基因序列，并将它们保存在 Genebank 中，研究人员可以很容易地在 Genebank 中找到同源序列完成系统学分析中的序列比对、引物设计等任务，这样可以有效地避免将假基因应用到系统学分析中。另一方面，虽然一些单拷贝的基因与 ITS 相比同塑性指数不高，但应用多个单拷贝基因作为分子标记同样可以得到合理的系统进化结果 (Alvarez and Wendel, 2003)。对 nrDNA 而言，应用更多的 nrDNA 序列 18S、26S、IGS 等序列进行统计、比较也可以得到可靠的结果 (Bailey 等, 2003)。栎属由于传统分类系统框架不清晰，其系统演化关系仍然是一个值得深入研究的命题。在 Muir (2001) 的研究中选择了两个分布于欧洲且两者种间杂交频繁的种，它们都是特例，是不是只有少数种类或易杂交的种类才会有 ITS 假基因，还是所有的栎属植物都有这一特点还需要做进一步的研究。目前，栎属的系统学研究中已应用到的分子标记有核糖体 rDNA 和叶绿体 DNA 中的几个片段，在今后的研究中应当尝试用更多的单拷贝的基因片段。另外，从 ITS 序列在栎属系统学研究的应用我们不难看出分子系统学在植物系统进化研究中既有优势又有不足，优势在于可以利用分子标记提供的信息很快地建立一个系统进化树以反映生物的系统进化关系；不足之处在于有些分子数据不具进化意义，不能提供有效的系统进化信息。有人曾探讨过应用基因序列分类的可行性 (Willem and Ferguson, 2002)，由 ITS 假基因的发现我们不难看出，应用单一的分子标记作为物种分类依据的风险，大量的实践已经证明传统的系统学理论在植物系统进化研究中的重要作用，只有二者的有机结合应用才能获得更有说服力的结果。

那些不能表达成蛋白质的基因曾被人们认为是存在于生物体内的“分子化石”，是生物在长期的演化进程中产生的一些无用的基因片段。Flam (1994) 认为这些基因也可能包含有一定的信息，只是人们并不完全知道这些基因真正的作用。Makalowski (2000) 认为那些重复的无功能的基因片段与和它们相邻近的基因相互作用而影响着生物的进化。在后来的研究中，Hirotsune 等 (2003) 发现与 Makorin1 同源的假基因 Makorin1-p1 能调控同源功能基因的表达。Lee (2003) 认为演化是双向的，一方面是在演化压力作用下，从已有的基因中产生新的基因；但不完整的基因制造过程也会产生有缺陷的基因片段，这些基因片段并没有永远失去作用，相反的演化压力会赋予它们新的任务。ITS 假基因是否也会有相似的作用呢？回答这一问题还需要做更多的研究。

〔参考文献〕

陈焕镛，黄成就，1998.《中国植物志》(22卷，壳斗科) [M]. 北京：科学出版社

邹喻苹，葛颂，王晓东，2001. 系统与进化植物学中的分子标记 [M]. 北京：科学出版社

Alvarez I, Wendel JF, 2003. Ribosomal ITS sequences and plant phylogenetic inference [J]. *Molecular Phylogenetics and Evolution*, **29**: 417—434

Axelord DI, 1983. Biogeography of oaks in the arcto-tertiary province [J]. *Ann Missouri Bot Gard*, **70**: 629—657

Bailey CD, Carr TG, Harris SA, *et al*, 2003. Characterization of angiosperm nrDNA polymorphyism, paralogy, and pseudogenes [J]. *Molecular Phylogenetics and Evolution*, **29**: 435—455

Buckler ES, Ippolito A, Holtsford TP, 1996. The evolution of ribosomal DNA: Divergent paralogous and phylogenetic implications [J]. *Genetics*, **145**: 821—832

Elias TS, 1971. The genera of Fagaceae in the southeastern United States [J]. *J Arnold Arbor*, **52**: 159—195

Flam F, 1994. Hints of a language in junk DNA [J]. *Science*, **266**: 1320

Govaerts R, Frodin DG, 1998. Word checklist and bibliography of Fagales [M]. Royal botanic garden, Kew

Hartmann S, Nason JD, Bhattacharya D, 2001. Extensive ribosomal DNA genic variation in the columnar cactus Lophocereus [J]. *J Mol Evol*, **53**: 124—134

Hirotsune S, Yoshida N, Chen A, *et al*, 2003. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene [J]. *Nature*, **423**: 91—96

Huang CJ (黄成就), Zhang YT (张永田), 1999. Flora of China (English version) Vol. (4)(Cycadaceae through Fagaceae) [M]. Science Press. Beijing & Missouri Botanical Garden Press. St. Louis. pp, 314—400

ISCID, 2004. Pseudogene [DB/OL]. ISCID Encyclopedia of Science and Philosophy. Retrieved December 16, 2004 from http://www.iscid org/encyclopedia/Pseudogene

Kita Y, Ito M, 2000. Nuclear ribosomal ITS sequences and phylogeny in East Asian Aconitum subgenus Aconitum (Ranunculaceae), with special reference to extensive polymorphism in individual plants [J]. *Plant Syst Evol*, **225**: 1—13

Lawrence, George HM, 1951. Taxonomy of Vascular Plants [M]. The MacMillan Company, New York

Lee JT, 2003. Complicity of gene and pseudogene [J]. *Nature*, **423**: 26—28

Makalowski W, 2000. Genomic scrap yard: genomes utilize all that junk [J]. *Gene*, **259**: 61—67

Manos PS, Doyel JJ, Nixon KC, 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* Subgenus *Quercus* (Fagaceae) [J]. *Molecular Phylogenetics and Evolution*, **12** (3): 333—349

Manos PS, Stanford AM, 2001. The historical biogeography of fagaceae: tracking the tertiary history of temperate and subtropical forests of the northern hemisphere [J]. *Int J Plant Sci*, **162** (6 Suppl): s77—s93

Manos PS, Zhou ZK, Cannon CH, 2001. Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution [J]. *Int J Plant Sci*, **162** (6): 1361—1379

Manos PS, Doyle JJ, Nixon KC, 1999. Phylogeny, biogeography and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae) [J]. *Molecular Phylogenetics and Evolution*, **12**: 333—349

Mayol M, Rossello JA, 2001. Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus* [J]. *Molecular Phylogenetics and Evolution*, **19**: 167—176

Muir G, Schlötterer C, 1999. Limitations to the phylogenetic use of ITS sequences in closely related species and populations– a case study in *Quercus petraea* (Matt.) Liebl. Chapter 11 in: Which DNA Marker for Which Purpose? [M]. Final Compendium of the Research Project: Development, optimization and validation of molecular tools for assessment of biodiversity in forest trees in the European Union DGXII Biotechnology FW IV Research Program Molecular Tools for Biodiversity. Gillet, EM (ed.). URL http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm

Muir G, Fleming CC, Schlötterer C, 2001. Tree divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. [J]. *Mol Biol Evol*, **18**: 112—119

Nixon KC, 1993. Infrageneric classification of *Quercus* (Fagaceae) and typification of sectional names [J]. *Ann Sci For Suppl* 1 (Paris), **50**: 25—34

Pu CX (普春霞), Zhou ZK (周浙昆), Luo Y (罗燕), 2002. A cladistic analysis of *Quercus* (Fagaceae) in China based on leaf epidermis and architecture [J]. *Acta Bot Yunnan* (云南植物研究), **24** (6): 689—698

Samuel R, Bachmair A, Jobst J, *et al*, 1998. ITS sequences from nuclear rDNA suggest phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American taxa of *Quercus* (Fagaceae) [J]. *Plant Syst Evol*, **211**: 129—139

Savolainen V, Chase MW, 2003. A decade of progress in plant molecular phylogenetics [J]. *Trends in Genetics*, **19** (12): 717—724

Soltis E. Douglas, Soltis S. Pamela, 2000. Contributions of plant molecular systematics to studies of molecular evolution [J]. *Plant Molecular Biology*, **42**: 45—47

Tian X (田欣), Li DZ (李德铢), 2002. Application of DNA sequences in plant phylogenetic study [J]. *Acta Bot Yunnan* (云南植物研究), **24** (2): 170—184

Willem J, Ferguson H, 2002. On the use of genetic divergence for identifying species [J]. *Biological Joural of the Linnean Society*, **75**: 509—516

Yang YW, Lai KN, Tai PY, *et al*, 1999. Molecular phylogenetics studies of Brassica, Rorippa, Arabidopsis, and allied genera based on the internal transcribed spacer region of 18s-25s rDNA [J]. *Molecular Phylogenetics and Evolution*, **13**: 455—462

Zhang DX, Hewitt GM, 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects [J]. *Molecular Ecology*, **12**: 563—584

Zhou ZK (周浙昆), 1992a. Drigin, phylogeny and dispersal of *Quercus* from China [J]. *Acta Bot Yunnan* (云南植物研究), **14** (3): 227—236

Zhou ZK (周浙昆), 1992b. A taxonomical revision of fossil evergreen Sclerophyllous oaks from China [J]. *Acta Bot Sin* (植物学报), **34** (12): 954—961

Zhou ZK (周浙昆), 1993. The fossil history of *Quercus* [J]. *Acta Bot Yunnan* (云南植物研究), **15** (1): 21—33

Zhou ZK (周浙昆), 1999. Fossils of the Fagaceae and their implication in systematics and biogeography [J]. *Acta Phytotax Sin* (植物分类学报), **37** (4): 369—385

Zoldos V, Papes D, Cerbah M, *et al*, 1999. Molecular-cytogenetic studies of ribosomal genes and heterochromatin reveal conserved genome organisation among 11 *Quercus* species [J]. *Theoretical and Applied Genetics*, **99**: 969—977